UNIVERSITY OF PENNSYLVANIA

AN EXPERIMENT IN MAP SCORING AND MENTAL IMAGERY TESTS

MARION ELDERTON

A THESIS

IN PSYCHOLOGY

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

PHILADELPHIA
1933

AN EXPERIMENT IN MAP SCORING AND MENTAL IMAGERY TESTS

A THESIS

IN PSYCHOLOGY

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF PENNSYLVANIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

MARION ELDERTON

Reprint from
The Journal of Applied Psychology, Vol. XVII, No. 4

PHILADELPHIA
1933

*30203

R. UNIV.
BIBLIOTHEEK
LEIDEN

AN EXPERIMENT IN MAP SCORING AND MENTAL IMAGERY TESTS

MARION ELDERTON
University of Pennsylvania

THE PROBLEM

This experiment was devised to determine possible relationships between differences in imaginal type, as determined by certain tests of imagery and the ability to draw a familiar outline map from memory. Map drawing and the use of outline maps play an important part in the teaching of economic geography. It has been generally assumed that a person endowed with good visual imagery would, for example, fare better in terms of an efficient map performance than one whose imaginal endowment did not include such imaginal ability. At the same time, it was also desired to know what relationship existed between the grades on maps as objectively given and those given by the method ordinarily used—*i. e.*, a subjective estimate of the teacher translated into quantitative terms. This relationship should be of interest to teachers of geography where any large proportion of the work consists of map drawing.

HISTORY

The literature on tests of imagery might be said to go back as far as Fechner, who published the introspective accounts of visual imagery of several well-known persons and gave an elaborate account of his own memory images and after-images. Sir Francis Galton in his "Inquiries into the Human Faculty and Its Development" (1883) describes the results of his questionnaires espe-

cially with what he called "the visualizing faculty." Charcot came to the question from the pathological aspect. These pioneers called attention to the striking differences between individuals in the nature and abundance of particular forms of imagery. Later work, however, was especially concerned with improvement in the investigating technique.[1] Fernald carried on an intensive experiment, using eleven trained subjects, the results of which were published in 1912.

The various types of tests which have been devised have been classified into objective, subjective and a combination of the two. Fernald defines an objective test as "one which furnishes results, either qualitative or quantitative, which are sufficient to indicate an individual's type without reference to his introspection." This does not mean that such tests may not have acquired this validity by the use of the introspective data originally but simply that it is not necessary to call for introspection at present.[2] In the case of children and persons unskilled in introspection, the desirability of such tests is obvious. The purely subjective test relies exclusively on the subject's introspection, while the combination subjective-objective tests use the result of one to check on the other.

While the tests used for this investigation combine the subjective with the objective the results will be considered entirely on the basis of their objective merits; brief discussion of the subjective aspects will be included as a matter of interest only. This will therefore eliminate at least for the present the attempt to classify the subjects according to ideational types, which is not essential in making the comparisons involved in the problem set

[1] Detailed critical discussion of techniques may be found by Angell, Colvin, Segal and Fernald. See bibliography at the end of this paper.

[2] Fernald, M. R., "The Diagnosis of Mental Imagery," *Psychological Review*, Monograph Supplement, 1912, Vol. 14, No. 1, P. 13.

forth. It is perhaps a recognition of the difficulties which beset the classification into distinct imagery types,[3] and of the tendency to search instead for individual differences in images. Fernald has aptly stated the problem when she says: "It seems significant of the trend of interest in the imagery problem as a whole that the question of 'imagery types' as such has for the most part moved into the background while the search for individual differences in images and for adequate methods of determining these has persisted. This may indicate a growing suspicion that such individual differences may be more profitably stated without the incubus of 'types'."[4]

A variety of methods has been used in studying individual differences in imagery. Under what might be termed "subjective" are mainly the introspective analyses by competent observers and the questionnaire; in the "objective" category mention might be made of the memory tests with their various modes of presentation

[3] In the historical development of concepts of imagery types, there has been the swing from the belief that an individual belongs in one of three distinct groups, "visual," "auditory," or "motor" or in a fourth, which combines the characteristics of the other three, the so-called "mixed" group, to that where it was maintained that it was practically impossible to discover persons belonging to the "pure" types, and that therefore most individuals belonged in a so-called "mixed" type group, thus minimizing differences which did exist. The exact meaning of the term "mixed" was also uncertain as to whether it applied to the various kinds of imagery used in combination for one task or that a change in imagery appeared with a change in conditions. Then there was a partial return to the earlier view of distinct types. It is agreed that there are marked differences in emphasis, even though individuals use more than one image form, and the terms "visual," "auditory," etc., may be retained if they are understood to mean the predominance of a particular form of imagery. The various difficulties in the way of this seemingly simple solution are brought out by Fernald in her study. (Pp. 18-20.)

[4] Fernald, M. R., *op. cit.*, pp. 20-21.

and reproduction, with their appeal to different sensory fields; word tests, where the conclusions are drawn from the content of the words. The tests falling into the "subjective-objective" class are similar to the memory tests except for the importance given to introspection. Association tests with introspection are sometimes included under this latter heading.

The introspective analyses by the competent observers, especially where they are long continued and where the training in introspection has been adequate, are considered among the best means of studying an individual's imagery, but these conditions severely limit their application to any large number of persons—eliminating most adults and all children. Moreover, this method is uncontrolled. The questionnaire, an outgrowth of the above, seems to have originated with Galton. One of its chief limitations has been expressed by Titchener, who says: "We must remember that the questionary can never transcend or go beyond the introspections of the individual correspondents."[5]

The letter squares fall in the category of memory methods which emphasize the order and arrangement of materials. In one instance the retention of a well-defined visual arrangement, without modification or in the order learned, is called for, and in another, the reproduction in a changed order. This method is sometimes referred to as "Binet's letter squares."

The assumption underlying the reproduction of the letters of a series in an order *different* from that in which it was learned is that one who visualizes the letters will find this change easier than one who relies on auditory imagery.

Binet, investigating the imagery of the lightning calculators, Diamandi and Inaudi, used the "different order"

[5] Titchener, E. B., *Experimental Psychology, Instructor's Manual*, Qual., p. 389.

reproduction of letters in squares and suggested caution in interpreting the objective results and concluded that "the introspection of the subject, in spite of its subjective character, ought always to be taken into consideration."

Titchener used the letter squares at the Cornell laboratory in a somewhat modified form and included introspection.

Angell considered the letter square test "one of the best . . . for bringing out verbal imagery, both visual and auditory motor."[6]

Of spelling tests, considered to overlap to some degree the letter square method, Angell made the statement that the former "ought in no event to be omitted" and that "spelling backward is (even) more useful than spelling forward."[7]

Commenting on oral spelling reversing, Fernald states: "In general, we seem justified in saying that the results indicate that the rapidity of this form of spelling is favored by the ability to summon clear, vivid and stable visual images of the words without a very great need for accessory factors. A poor record, however, need not indicate an absence of visual verbal images, though it may be due to that. It may also indicate that the visual image will not stand alone, but needs other factors of the same character as those used by non-visual subjects, or that the visual image is not of a suitable character for this operation: *i. e.*, it may be of word-whole in which the letters do not stand out clearly, or it may constantly change form or disappear during the operation."[8] It is considered a partially objective test lending confirmation to the introspective reports of the subjects.

The written spelling reversing offers imagery cues

[6] Angell, J. R., "Methods for the Determination of Mental Imagery," *Psychological Review*, Monograph Supplement, 1910, Vol. 13, No. 53, p. 92.

[7] *Ibid*, p. 93.

[8] Fernald, M. R., *op cit.*, pp. 54-55.

similar to the oral, although the queer appearance of the written word tends to act as a real distraction[9] in many cases.

The foregoing will give a brief idea of the background of some of the more important imagery tests and how they have developed.

Method of the Investigation

In view of what has already been said with reference to the difficulty of classification of individuals into distinct "imagery types" and to the lack of really adequate technique to make such classification on the basis of strictly objective tests, no such attempt will be made here. The treatment of the test results will be as objective as possible, even though introspection was called for from the beginning in case it might have been of value in interpreting the individual differences. Observations of the experimenter were also noted wherever possible for whatever light they might throw on the method used by the subject. No attempt, however, will be made to offer a definite diagnosis of the subject's imagery type, partly because of some subjective factors which would thereby be injected, and partly because the purpose of the study can be accomplished without it.

A series of tests, to be described, was developed and administered to over 400 university students (most of whom were Freshmen) who were enrolled in a required

[9] The so-called "distraction principle" as a method has sometimes been used in imagery tests. It is brought out somewhat in the written spelling performance. The assumption is that "a sensory activity or an imagery process of any sort interferes with the simultaneous arousal of imagery of the same kind. Thus, sounds conflict with auditory imagery, sights with visual, movements with motor, etc. Any diminution of efficiency under conditions of distraction is taken as an evidence, therefore, that the given disturbance is allied to a favored form of imagery. The extent of the interference is considered a measure of the dominance of this imagery." (Fernald, p. 10.)

Freshman course in economic geography of the United States. In this course outline maps of the United States were constantly being used by the students as graphic aids in studying the various economic-geographical relationships. Outline maps of a certain size were usually distributed for this purpose, but occasionally the map was drawn free-hand by the student. For purposes of the experiment the drawing of an outline map of the United States was included as a part of the final examination in the course. These were graded on the basis of from 1 to 10 by the student's individual instructor—10 being the grade for the most satisfactory map.

Realizing the difficulty of securing a reliable criterion on this basis for an objective scoring of these maps they were discarded and another map was later secured from most of these students. Some of the students, however, had left the university in the meantime. For the second map each student was given a sheet of cross-section paper (5 blocks to the inch), proportioned for a correct map of the size which had been customarily used by them, with the mimeographed instructions in the center of the sheet as follows:—

> Please draw an *OUTLINE* map of the United States the maximum size that can be drawn on this sheet. Don't attempt to put in all the kinks of minor bays or capes. Remember that it is to be as large as you can possibly get on the paper. You may draw in the Great Lakes if it is easier for you to get the outline of that section in that way.

The student, of course, was not informed concerning the correctness of the proportions of the sheet although this could have been implied from the instructions. A stencil for checking the accuracy of the map was used

in scoring and the number of block deviations between the student's outline of the map and that of the stencil was the map score. Then, in order to secure a subjective (or an instructor's) score on the *same set of maps*, each map was traced by the experimenter on a plain sheet, 8½x11 inches, and the resulting carbon copy was turned over to an instructor to grade, on the basis of from 1 to 10,—10 being the best score and 1 the poorest. It was not originally intended that the instructor should use any correct sample as a guide or aid in his grading. However, due to his desire to do a careful job it was discovered that a correct sample had been employed. Because of this unintentional procedure the results are even more striking. The person scoring the maps on the subjective basis was informed as to the instructions under which the original maps had been drawn. In most instances, because of its having been transferred to a much larger sheet, the traced map had the four corner boundaries indicated to show the size of the original cross-section sheet.

Turning now to the testing materials the following problems were included in the original series:—

1. A mimeographed questionnaire, adapted from the Galton Breakfast Table Blank and the Titchener questionnaire, was used in an attempt to get some information regarding the subject's endowment and perhaps some clue as to the predominant type of imagery used.
2. Twenty (20) words presented orally to be reversed orally by the subject on which time and errors were recorded. Time was taken to tenth of a second. The observer used a mimeographed list of the oral words, in reverse order, which were easily checked as the subject spelled aloud. For example, omissions were circled, extra letters inserted. transpositions indicated, etc.
3. Twenty (20) words different from the above twenty to be written in reverse order upon the oral presentation of the stimulus word. Time was taken and recorded, as above, and errors were later checked.

4. A letter square (known as Letter Square "A") consisting of nine (9) letters arranged in three rows, and in such a way as to minimize the possibilities for association. The oral response of the subject, learning time and reproduction time in the so-called "normal" order were recorded.
5. The above series of letters was then used, with instructions to reproduce orally in the so-called "spiral" order (known as Letter Square "B"). Letters reproduced and reproduction time were recorded.
6. A letter square (known as Letter Square "C") consisting of a different set of nine (9) letters, arranged in three rows, was learned with learning time constant, and the letters given and reproduction time in the "normal" order were recorded.
7. Memory span for digits, presented visually and orally, for both forward and reverse spans.

In addition the subject's introspection, whenever possible, was secured on the Letter Square "B" ("spiral" reproduction) and on the oral and written word reversing. Observations by the examiner of the subject's overt behavior with respect to such items as lip movements, pronouncing aloud, using fingers, closing eyes, etc., were recorded where noted.

As the study progressed it became necessary to cut down somewhat on the amount of time which could be devoted to the testing of each subject. The following eliminations were subsequently effected:—

1. The oral word list was reduced by one-half,—the first ten words being retained.
2. The written word list was also reduced by one-half,—by ten words.
3. The memory span tests for digits.
4. The questionnaire.[10]

[10] The questionnaire was discarded in the treatment of results primarily because the information thus secured was of a highly subjective nature. It may be said, however, to have been of some value as it stimulated the student inexperienced in introspection to think about it. Since the questionnaire was usually distributed

All the tests were given as individual tests except the questionnaire.

The oral spelling word list was arranged so that a shorter word,—of four or five letters—alternated with the longer words in most cases. Thus in the ten-word oral list, there were two four-letter words and two five-letter words. This was done in order to ease up somewhat on the strain of spelling the longer words.

A similar arrangement held for the written list.

For the purposes of this experiment the scores on the short words have been disregarded. The score on the remaining longer words was then computed, taking time and errors into account. For example, remaining on the oral list there were six so-called long words, *i. e.*, ranging from nine to twelve letters. To the sum of the actual times taken was added a penalty of twenty (20) for each error,[11] and this figure then subtracted from 700. For example, let us say that scores were made of 15.5″, 10.2″, 7.1″, 9.2″, 18.0″, 18.0″ on the six oral words respectively, with one error,—which would result in a total score on that test of 602 (700–[78+20]). On the other hand, if say, ten errors had been made by the subject the penalty of 200 (10×20) would be added to the reproduction time (78 seconds) before subtracting from 700, resulting in this case in a score of 422. Hence a high score, *e. g.*, 602, would indicate rapid reproduction and few or no errors or a combination of speed and accuracy.

The list of words to be reversed and reproduced in written form contained five short words (four of four

to a class a week or more before the individuals in it reported for their tests, it may have helped to give the student some idea of his own imagery processes and perhaps his later introspective reportings were better than they otherwise might have been. It was not uncommon for a student to remark, "I had never thought of how I remembered things before, and since filling out the questionnaire, I've been checking up on myself to see what I do."

[11] A transposition of letters, omissions, extra letters, etc., were penalized by adding 20 for each such error.

letters each and one of five letters) and five longer words —two nine-letter and three ten-letter words. The scoring followed was similar to that for the words presented and reversed orally.

The words chosen were ones with which a college Freshman would ordinarily not have difficulty. There were no "trick" words and allowance was made for a certain type of misspelling by not penalizing the subject, if he reversed a syllable, as for example, "ence" even if the syllable to be absolutely correct was "ance." This seemed a fair thing to do since our interest was not primarily in the subject's ability to spell words correctly as such but on the reversing of what he believed to be the correct sequence of the letters in the word presented to him.

At the conclusion of the written list, which was given subsequent to the oral, an attempt was made to secure the introspection on the spelling operations.

In some cases it was hardly necessary to ask for an introspection because observance of the subject revealed the moving lips and slight throat movements used in spelling over letters or pronouncing syllables together with the halting reproduction of one, two, or three letters at a time. Sometimes spelling forward aloud was noted.

The letter square material for our experiment was divided into three parts. Letter Square "A" consisted of a card containing a square divided into nine spaces, each space containing one letter. For this series the subject was handed the card with the instructions to learn the letters across from left to right and then to turn the card over face down, when he had learned them, and to wait until he was asked before telling the examiner what the letters were.

The criterion of learning here was the honest attempt to learn the letters in order to reproduce them orally in the "normal" order when asked for by the examiner.

Approximately five seconds elapsed between the time the card was turned face down and the letters called for. The learning time, reproduction time and letters given were then recorded.

The next part known as Letter Square "B" involved this same series of letters. However, in this case the subject was asked to reproduce the letters in the "spiral" order, this order being indicated by a blank letter square with a red arrow drawn through the squares to indicate the order wanted. No further learning was permitted and the examiner entered the letters as given by the subject and recorded the time.

Letter Square "C" involving a different series of nine letters was given with the learning time kept constant, *i. e.*, 18 seconds. The subject was informed that in this instance a similar series of letters would be shown him to be learned in the order indicated (*i. e.*, the so-called "normal" order), but that there was a time limit on learning and that the examiner would turn the card over when that time was up. The subject was also assured that he would not be asked to reproduce the letters in any other order than the one learned. This was done to prevent him from anticipating a different order and perhaps thereby "wasting" his limited time on a problem not to be included. The time of reproduction was recorded for this series.

Introspection was usually requested on Letter Square "B" (the spiral order performance). It was often necessary to ask the subject when attempting to secure this introspection how he remembered the letters up the right row of the series as he proceeded around from the bottom.

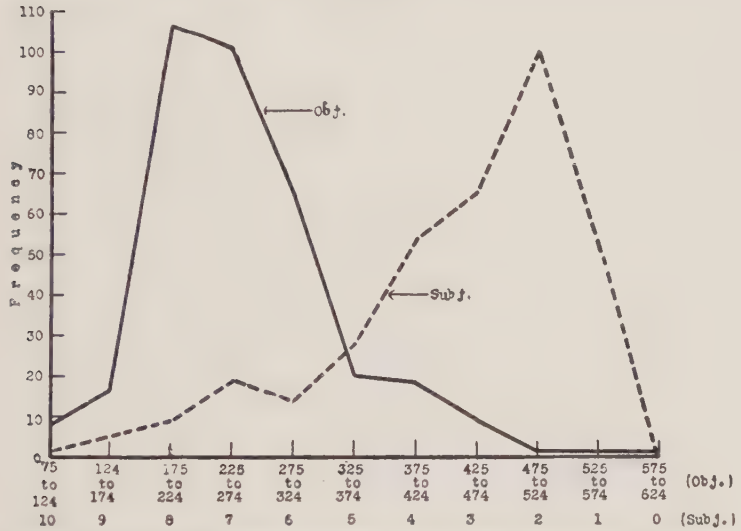
Results

The results on the map scores will be given first because the data on the imagery tests are treated on the basis of grouping by map scores, which grouping will be explained and discussed later.

Figure 1 shows the distribution of the map scores both on the stencil grading and on the instructor's grading bases with the "good" scores on both laid out from the left hand side. For convenience in referring to these two sets of scores they will be designated as "objective" and "subjective" respectively.

FIGURE I

Distribution of Map Scores—Objective and Subjective Scoring Bases



(Note: On Obj. Scores—low scores are good; on Subj. Scores—high scores are good)

TABLE I

Map Scores—Objective and Subjective Bases

	Objective	Subjective
Median	242.3	3.32
Avg.	250.5±3.9	3.29±.1
S. D.	71.5	1.9
Q	44.4	1.2
V	28.5	57.5
Skewness	+.34	—.05

The combining of these results in one graph brings out clearly the significant relationship. It should be remembered that low objective map scores are good ones because they represent less deviation from the correct stencil outline, whereas on the subjectively scored maps the higher scores represent the better performance. The distribution of the objective scores is piled up at the "good" end while the line indicating the distribution of subjective map scores piles up at the "low" or poor end.

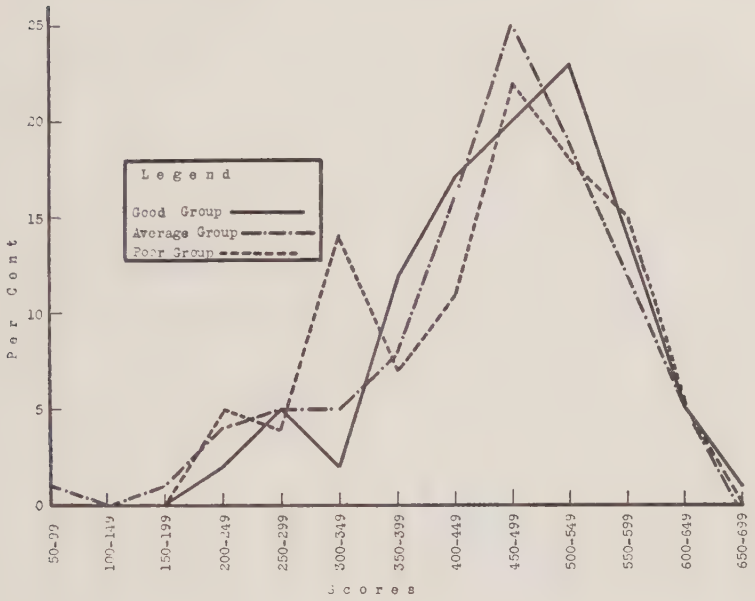
The coefficient of correlation (Pearson) computed for these two series is –.33 with a PE of .03. (See footnote, p. 404.)

We are now interested to discover how the map drawers fared in the various imagery tests which were administered. In order to determine this the subjects were divided on the basis of their objective map scores into three groups, good, average and poor, using the upper quartile, the two middle quartiles, and the lowest quartile respectively. For example, those receiving a score up to and including 206 on their map were classified in the "good" group; those receiving from 207 to 287 in the "average" group, and those receiving 288 and above in the "poor" group, *i. e.*, the greater the number of block deviations, the poorer the map and the higher, numerically, the score.

The significant measures for the various tests according to these groupings are to be found in the following tables and figures. Graphic presentation is worked out on a per cent frequency basis in order to make the results comparable. In view of the rather general negative results which characterized the data, the several charts included give a representative picture of the entire series.

FIGURE II

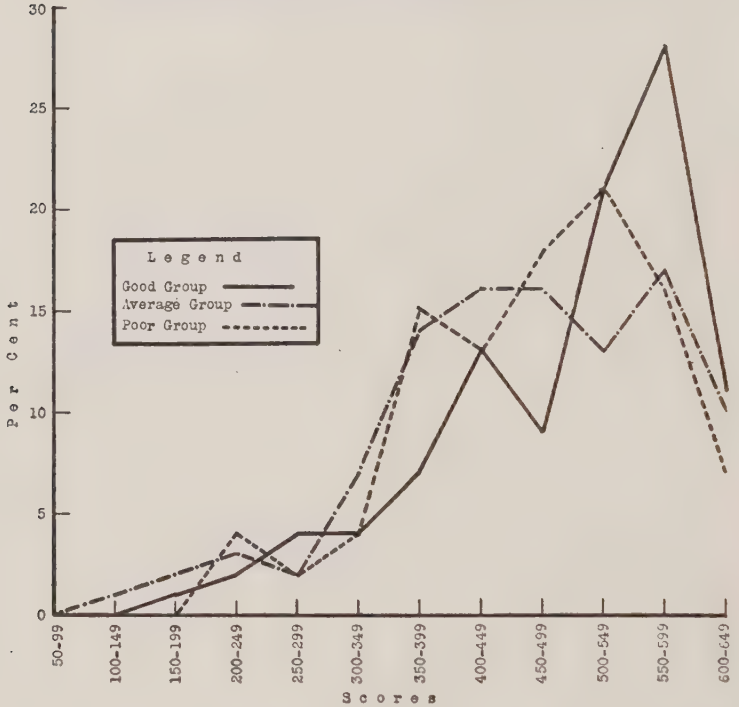
Distribution of Scores on Oral Words (Long) for the Three Groups Divided on Objective Map Score Basis



	Good Group	Avg. Group	Poor Group
Avg.	471±10.0	455.5±7.9	454.0±11.4
S. D.	93.5	103.0	105.0
Q	62.8	62.0	84.6
V	19.8	22.6	23.1
Sk	—.32	—.43	—.53

FIGURE III

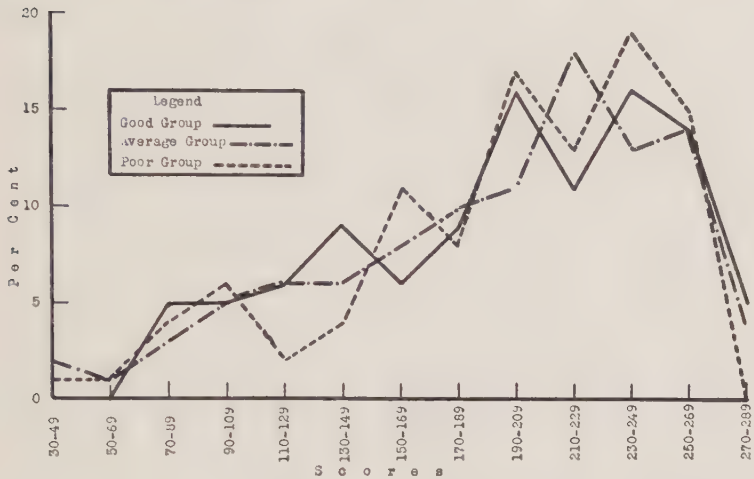
Distribution of Scores on Written Words (Long) for the Three Groups Divided on Objective Map Score Basis



	Good Group	Avg. Group	Poor Group
Avg.	494.5±11.4	459.5±8.9	470.5±1.0
S. D.	105.0	115.5	97.5
Q	73.0	85.0	74.0
V.	21.2	25.1	20.7
Sk	—.83	—.20	—.38

FIGURE IV

Distribution of Scores on Letter Square "B" (Spiral) for the Three Groups Divided on Objective Map Score Basis



	Good Group	Avg. Group	Poor Group
Avg.	196.4±5.8	193.2±4.5	195.8±5.8
S. D.	54.6	57.6	54.4
Q	43.7	41.3	37.5
V	27.8	29.8	27.8
Sk	—.43	—.67	—.56

TABLE II

Letter Square "C"—Objective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	219.3±5.8	227.3±4.1	224.3±5.9
S. D.	50.8	49.8	49.8
Q	41.4	41.1	37.2
V	23.1	21.9	22.2
Skewness	+.11	—.30	—.42

Since the scoring of maps is ordinarily not done on such an objective basis but on the subjective one mentioned earlier the subjects were re-classified into three

similar groups, good, average and poor, but this time according to their "subjective" map score. Their test results on the imagery material follow.

TABLE III

Oral Words (Long)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	460.5±11.1	466.0±7.4	446.0±10.9
S. D.	101.5	97.5	99.5
Q	60.8	67.9	73.6
V	22.0	20.9	22.3
Skewness	—.43	—.26	—.71

TABLE IV

Written Words (Long)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	484.5±11.9	477.5±7.8	448.0±13.0
S. D.	110.0	101.0	119.0
Q	87.0	74.4	85.4
V	22.7	21.2	26.6
Skewness	—.49	—.40	—.34

TABLE V

Letter Square "B" (Spiral)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	201.6±5.6	194.6±4.2	179.8±6.6
S. D.	51.6	55.4	60.2
Q	38.6	35.9	50.0
V	25.6	28.4	33.5
Skewness	—.63	—.79	—.64

TABLE VI

Letter Square "C" (Normal)—Subjective Map Basis

	Good Group	Avg. Group	Poor Group
Avg.	217.5±6.0	224.0±4.2	233.0±5.2
S. D.	52.8	50.5	45.0
Q	45.9	41.4	36.5
V	24.3	22.5	19.3
Skewness	—.15	—.18	—.37

Discussion

The relationship between two methods of grading outline maps is brought out in the results of the map scores in this experiment. Figure 1 (solid line) shows the distribution of the objectively scored maps to be skewed to the right (positive), which in this case happens to be the high end numerically but the poor end qualitatively, since the larger the number of block deviations, the poorer the map. A check on the last two cases reveals a real failure to follow the instructions, in both cases a rather well-proportioned map is drawn on a much smaller scale than the standard instructions called for. Adjusting a stencil for a smaller map, these two scores would be quite good. If these two cases were discarded the skewness for the distribution is changed to +.31 as against +.34. The average is also raised qualitatively (hence lowered quantitatively) to 248.7 as against 250.5.

If the criticism be made that an otherwise good map is penalized by the use of the stencil, because of the necessity for the stencil to be definitely placed at certain points, the answer seems to be that in the ultimate deviation there would be no benefit to the score. For example, it might be claimed that in a particular map the eastern coast line is quite accurate but started over to the left too far, and by placing the stencil in its correct place to the right the "goodness" of the east coast line is not properly evaluated. Suppose then the stencil were shifted to the left to "fit" the eastern coast, then the map drawer would lose out at the other side due to the fact that the left side of the sheet limits the subject from proportionately extending the western outline and so the points gained by shifting the stencil to the left to take advantage of a good Atlantic Coast line would be offset at the other side by the fore-shortening of the Pacific Coast line because of the limitations of the paper on which the map was drawn.

It may be asked, why bother to grade maps so care-

fully when a conscientious effort on the part of the teacher can do a similar job with much less time involved? As far as the time element is concerned there is no argument although the additional time involved with a stencil is not great, but when a *similar* job is desired there may be some room for doubt. However, on a task that is expressed in such an objective manner as the outline of a map one might logically expect a rather close relationship to result from the two methods.

A most careful job was done on the marking of the duplicate set of maps with the aid of a carefully prepared copy of the size of the original map kept before the instructor and with the four corners indicated of the boundaries of the cross-section sheet.

The distribution of these scores, known as the "subjective" basis scores, is also shown in Figure I (broken line) and arranged with the best scores, qualitatively, at the left to correspond to the arrangement in the other map series. The distribution here is skewed (–.05) in the opposite direction. The piling up is at the low score (*i. e.*, poor map) end, and the mode is at 2. Superimposing the two curves on the same figure makes the lack of similarity apparent.

When the coefficient of correlation (Pearson) was computed it was found to be –.33, with a PE of .03. (See footnote, p. 404). This relationship while slight is statistically significant, being eleven times its probable error. Such a result is surprising in view of the careful, conscientious job which the writer knows was done on this set of maps. This is another example of relative unreliability of subjective teachers' grades as compared with objective criteria.

Imagery Tests and the Objective Map Results

In order to make comparisons between the results on the imagery tests and the map scores it will be recalled that the subjects were divided into three groups—the

good, average and poor—on the basis of the map score. This comparison is made first on the objectively scored maps. Spelling words backwards and letter square tests constituted the imagery tests used in this experiment.

Tests on backward spelling have been used by a number of investigators although Angell states that they seem to have been used less than one might anticipate. He suggested that they are an admirable form of test to use in conjunction with tests on descriptions inasmuch as it brings into the foreground peculiarities of verbal imagery as compared with the imagery of objects, and that this virtue constitutes its main limitation. It would seem, however, that if one does not interpret the term "verbal" too literally this test also seems to combine the "object" or concrete category with it. The task of spelling a fairly long word backward sets up the combination of letters *as an object*,—they are seen in different kinds of letters,—large, small, printed, written, fancy, black on white, and in various spacings, etc. The short four- and five-letter words scarcely presented any problem and the response was more or less automatic, so that they were excluded in treating the results. They did, however, fulfill a real need in furnishing relaxation and encouragement, coming as they did between longer words. The latter to many presented a real problem and there was little opportunity for the performance to become an automatism.

One of the interesting things brought out by the introspection and which seems not to have been stressed by investigators of imagery was the importance of the motor element, even for those whose visual images were clear, easily stimulated, and relatively lasting. In nearly every case the subject admitted that before any helpful cue for reversing was obtained some pronouncing forward to himself or a spelling of the word forward to himself was the first response. This would call up either a visual image of the whole word or parts of the word,

of syllables or one or two letters at a time and if the image did not last sufficiently long, it was continuously revived and reinforced by auditory and/or motor images. Except in cases of especially good visualization, where the word could be set up visually and read off backwards at an even rapid rate, there was a constant spelling forward (or pronouncing) to the place where the subject had left off reversing. Cases were found where audible spelling and pronouncing had to be resorted to in order to do the test. It was common testimony that when the letters disappeared or faded, the subject would spell forward or pronounce again and again. Shaking the head to emphasize syllables and counting on the fingers often accompanied it. Sometimes in the struggle to hold the images the subject would become confused and not remember accurately "where he was." This would sometimes involve the insertion of extra letters or syllables or of transposition of letters. A number of records, a very small proportion, indicated the rapidity and accuracy of those whose introspections bore testimony to clear, lasting visual images. An even smaller number on this test whose auditory ability seemed highly efficient compared almost as favorably as did some of the former. In other instances one could see the lips actually moving in efforts to retain enough of the image of the word to give back, haltingly, a letter or two at a time. If you think it is an easy matter, especially if you happen not to be highly endowed with clear visual images, try to spell backward "Philadelphia" or some other equally long word. Where the vocal-auditory and visual cues were efficiently combined in the subjects a fairly creditable performance resulted.

If we take now the three groups divided on the basis of the map drawn (and objectively scored) and examine the results obtained on the oral spelling test we find (see table below Figure II) that there is little differentiation. It will be remembered that a high numerical score on the

oral word test is a good score. There is a slight difference in favor of the "good" group with an average score of 471, but there is practically no difference between the means of the "average" and "poor" groups, 455.5 and 454.0 respectively. The coefficient of variation (Pearson) for the entire group is 22.3 as compared with 19.8, 22.6 and 23.1 for the separate groupings. There is a negative skewness which increases from the good to poor groups.

The graphic representation of these data (Figure II) shows clearly the essential similarity of these groups.

The scores on the written words are shown graphically in Figure III. The written spelling operation seemed to be somewhat easier for many of the subjects because as they themselves expressed it they "could tell where they were" more readily, and did not need to spend so much time remembering where they had stopped on the reversing operation. On the other hand the queer appearance of the words written this way seemed to confuse and distract in some instances. This was brought out particularly by the introspection of those who had claimed good visual images, who insisted that the appearance of the writing "knocked the word completely out of" their minds, and it was only after they hit upon the method of writing with the eyes closed or of looking away while writing, that this distraction was removed. This testimony seems to lend some weight to the so-called "distraction principle" mentioned earlier. The effect was certainly more noticeable with that small proportion of the group who claimed great ease of visualization.

On the whole the written scores were better than the oral regardless of the extra time which was presumably involved in the manual operation of writing.

This seems to be somewhat at variance with Fernald's findings on the written spelling tests. In summarizing this section of her study she states: "In general, it is evident that the procedure is more difficult than the preceding (oral), chiefly because the peculiar appearance is

sufficiently disturbing to more than counteract any advantage that might be expected from the sight of the letters already written."[12] It can hardly be accounted for on the basis of different words, since the five longer words retained on the written spelling list were taken from the Fernald list. It might, however, be partly due to the fact that the average length of the written words was 9.8 as compared with 10.7 letters for the oral.

In the results for this test the best group continued to show the highest average, but the difference between it and others is not statistically significant. The low group is the least variable, although there is little difference in this respect when all are considered. All three curves are skewed at the high end, but the best group shows the greatest amount, as might be expected. Confirmation, therefore, on the basis of written scores of a real difference among the groups according to map drawing ability is lacking here.

In the letter square test spatial localization of the items which are used in the squares is considered as a help in contrast to the so-called "distraction principle." Angell states that it is one of the best in the list (of those his committee was evaluating) for bringing out verbal imagery, both visual and auditory-motor.

In the performance of Letter Square "B" it may be recalled the subject reproduced a series of nine letters in a spiral order, having learned them in a so-called normal order and without knowing in advance that a different order would be called for. The scoring here penalizes for errors and takes account of speed of reproduction as well as of learning, for it includes the sum of the reproduction and learning times plus the penalty of 20 for each error,—this whole sum subtracted from 300. A high score, therefore, is a good score. An error is defined as an omission, a letter substituted or not in the list or a letter not in its correct "box."

[12] Fernald, M. R., *op cit.*, p. 57.

In regard to learning time, this was a variable factor since the criterion was simply the honest attempt to learn the series so that it could be reproduced in the normal order.

A changed order of reproduction for a good visualizer, all other things being equal, has ordinarily been assumed to be a relatively easy task and one should therefore expect to find a differentiation among the three groups, assuming further, of course, that excellence of map drawing represented some relationship to the individual's dominant form of imagery.

Figure IV reveals practically no differentiation whatever and the results are almost identical with those for the group as a whole. If the headings were omitted from the columns one would hardly realize that three supposedly distinct groups were being compared.

The part played by association is difficult to evaluate. While the letters chosen were supposed to minimize this aspect of the test, it was surprising what ingenuity was resorted to in an effort to remember the letters. The most common type of association, however, was that made with initials of names. As far as the introspection was concerned it seemed that the saying over to oneself of the entire line to get the needed letter was most frequent. This process might be assumed to be accomplished quickly without any great amount of visual cues if the original learning had been carefully done. It was the rare subject to whom all the letters appeared simultaneously and remained long enough to enable him to merely read them off in the order wanted. In several instances subjects reported that the three letters on a line at a time "flashed into mind" and gave sufficient usable cues to reproduce without any further effort. To be sure these individuals were the exceptions. Most subjects, however, reported auditory-motor with visual in varying degrees. These tests as a whole tended toward

the appearance of visual memory images even though somewhat vague in many cases. It appears that the articulatory image oftentimes gives the cue for the reinstatement of the visual image when the latter has been lost.

In order to equalize somewhat the difference between the slow and rapid learners the learning time on Letter Square "C" was kept constant at 18 seconds. This amount of time set arbitrarily turned out to be about half of the average time taken to learn Letter Square "A"—computed from those cases where the reproduction was perfect in the so-called normal order. On Letter Square "C," the subject was definitely assured that he would not be asked for a reproduction in any but the order in which he had learned the letters.

Table II gives the significant results on the scores for this series. The score was computed by subtracting the sum of the reproduction time and 20 times each error from 300. It will be noted that the range of scores in this series is much narrower than in Letter Square "B." The performance of a task in an accustomed manner, *i. e.*, in the normal order, does not appear to make for any great variation.

The negative picture, however, is continued here so far as differences among the high, middle and low map groups are concerned. The middle group seems very slightly better than the other two.

Due to the part played by memory in learning the squares and the difficulty of controlling overlearning or underlearning, it would seem that the advantage in tests of this kind might be in favor of the spelling tests. Here, if the subject knows the word or if it is not too difficult, he has an opportunity to use the imagery best suited to him. However, it cannot be assumed that he will always use this procedure.

Imagery Tests and the Subjective Map Results

There would be no point in going further and making a similar analysis of the imagery test scores, re-sorting into three groups by the map scores graded on the subjective basis, were it not for the fact that the latter method of grading maps is practically the only one used, and had the other procedure not been tried out the customary method might have been considered the logical basis on which to have made the sortings.

The groups here were divided similarly to the previous three, *i. e.*, the highest quartile (good), the two middle quartiles (average) and the lowest quartile (poor). A consideration of scores on oral words spelled backwards shows that the mode is the only measure in which the good group can claim a seeming superiority, 525 against 475 for the other two groups. The mean of the average group is slightly better than that of the good group, although not significantly so. The poor group average, however, is definitely lower than the others, 446 as against 460.5 (good) and 466 (average). But when the variation is considered the groups present no clear-cut differentiation. (See Table III.)

From Table IV (written scores for words spelled backward) we get some differentiation. The averages for the groups show a decreasing tendency from the good to the poor group—484.5, 477.5 and 448—which is what one would be led to expect—but in the other measures the average group varies least from the central tendency so that here again, when the whole picture is taken into consideration, we cannot argue for any clear-cut evidence that the subject's ability displayed by his map work is closely related to his performance on this particular test.

An analysis of Tables V and VI present two nearly opposite pictures. In the former with respect to Letter Square "B," we have practically the only data which give

a consistent advantage to the good group, a lesser one to the average or middle group, and the poorest showing to the third group.

Letter Square "C" results on the other hand are almost the reverse, with the poor group carrying off the laurels for superiority, such as they are.

Referring back to the actual map drawing and the most adequate way to execute a task such as this from memory, it has been claimed that an individual who is a ready visualizer, who can "see clearly in his mind" the whole outline at a time or even parts of it, should be able to translate that picture easily and effectively into an accurate outline. On the other hand it is conceivable (and demonstrable) that a person totally deficient in visual imagery could make a very creditable showing on such a task, with the imagery cues which come from the kinaesthetic field predominating. Just consider for the moment that you have been asked to draw such a map as was called for in this experiment and suppose also that you do *not* have a very clear picture of it in your mind. Do you find yourself using auditory-motor verbal aids in locating the various points? For example, do you find yourself saying that the New England States come down like this, Cape Cod juts off here, New Jersey turns in thus, the Gulf of Mexico comes up and goes around like this, etc.? Then the instant you make a line on your sheet of paper you begin a comparison of the actual with the memorial or imaginal, shifting quickly back and forth, checking the image, and getting the cue for the next movement. It seems if one performs the task slowly and carefully that the auditory and vocomotor verbal aids are likely to be experienced and prove helpful. Whatever ability the map drawers in the best group had in the experiment it showed little differentiation when set up along side the other two groups' results on the imagery tests.

Conclusions

1. The most striking conclusion which appears refers to the method of grading the maps. From the results of this experiment, a slight although significant relationship is shown to exist between subjective and objective map scores, represented by the correlation of only –.33 (PE .03).* There is an indication that the instructor who did the scoring tended to keep the general shape of the map in mind, as indicated by this positive relationship, which of course does not have the accuracy made possible by the actual objective measurement.

2. There are no significant differences in correlation between map scores and tests of imagery among the three groups on either basis of sorting. The closest approach to consistent (but not statistically significant) differences is found in the results of the Letter Square "B" (subjective basis) where the good group is somewhat superior to the average, and the latter to the poor group.

3. A consistent inverse relationship is found to exist among the groups for Letter Square "C" (subjective basis), but it is not statistically significant.

4. Whatever abilities are involved in the drawing of the map they are not significantly related to those involved in the imagery tests as administered and scored in this experiment. The former may be much more complex and varied than was assumed.

5. At the university level, regardless of the advantages which might be obtained at lower levels, the actual drawing of maps might be dispensed with in view of the

* This coefficient is preceded by a minus sign because high scores represent good performance on one test and low scores good performance on the other.

low degree of correlation between the methods of scoring the maps.

6. It might be interesting if a satisfactory technique were evolved for classifying the subjects into four groups on the basis of the introspection and combined score computed for all the tests into a so-called (1) good visual, (2) intermediate (vis.–aud.–mot.), (3) poor visual and (4) good auditory. Then a comparison could be made with the maps which would represent an opposite approach to the one which has just been made. This kind of classification necessarily emphasizes the subjective aspect but the introspection would have to be used as a check for the objective test record.

7. The paper presents new objective methods for both the scoring of drawn maps and for the combination of times and errors in a single score for the tests of mental imagery.

Bibliography

(Note: Only a few selected references will be included here, chosen primarily because of their historical significance. The bibliography contained in Fernald's study contains 118 references, which may be consulted should more material be desired.)

1. Angell, J. R., "Methods for the Determination of Mental Imagery," *Psychological Review,* Monograph Supplement, 1919, Vol. 13, No. 53, 61-107.

2. Bell, A., and Muckenhoupt, L., "Comparison of Methods for the Determination of Ideational Type," *American Journal of Psychology,* 1906, 17, 121-126.

3. Binet, A., *Psychologie des grands calculateurs et joueurs d'échecs.* Paris, 1894.

4. Charcot, J. M., *Leçons sur les maladies du système nerveux. Oeuvres Completes,* 1886-1890, 3, 154-193.

5. Colvin, S. S., "Methods of Determining Ideational Types," *Psychological Bulletin,* 1909, 6, 223-237.

6. Fechner, G. T., *Elemente der Psychophysik,* 2, 469 ff.

7. Fernald, M. R., "The Diagnosis of Mental Imagery," *Psychological Review,* Monograph Supplement, 1912, Vol. 14, No. 1.

8. Galton, F., *Inquiries into the Human Faculties.* London, 1883.

9. Galton, F., *Statistics of Mental Imagery. Mind,* 1880, 5, 301.

10. Segal, J., "Ueber den Reproduktionstypus und das Reproduzieren von Vorstellungen," *Arch. für das ges. Psychol.,* 1908, 12, 124-236.

11. Titchener, E. B., *Experimental Psychology, Student's Manual.,* Qual., 195-200; *Instructor's Manual,* Qual., 387-402.